

PAPER

Global Nonlinear Optimization Based on Wave Function and Wave Coefficient Equation

Hideki SATOH[†], *Member*

SUMMARY A method was developed for deriving the approximate global optimum of a nonlinear objective function with multiple local optimums. The objective function is expanded into a linear wave coefficient equation, so the problem of maximizing the objective function is reduced to that of maximizing a quadratic function with respect to the wave coefficients. Because a wave function expressed by the wave coefficients is used in the algorithm for maximizing the quadratic function, the algorithm is equivalent to a full search algorithm, i.e., one that searches in parallel for the global optimum in the whole domain of definition. Therefore, the global optimum is always derived. The method was evaluated for various objective functions, and computer simulation showed that a good approximation of the global optimum for each objective function can always be obtained.

key words: *nonlinear, global optimization, wave function, quantum computing*

1. Introduction

Conventional global optimization methods are classified into random search methods and gradient methods. The former randomly evaluate an objective function to search for the global optimum. That is, they have to evaluate the objective function over the entire search space in principal. They thus need much computation time to find the global optimum; otherwise we cannot confirm whether the solution obtained is the global optimum. The latter use the derivative of an objective function to asymptotically search for the global optimum. The objective function thus must be differentiable, and the global optimum cannot always be derived although less computation time is needed.

One of the most serious problems in deriving the global optimum is avoiding falling into a local optimum because it is quite possible for all the methods expect the full search method to fall into a local optimum. Tunneling algorithms [1], [2], [3] first search for a local optimum using a gradient method and then search for a better local optimum using a tunneling method starting from the local optimum previously obtained. The complex dynamics of a chaotic attractor is applied to various optimization methods to avoid being trapped in a local optimum [4]. Hopfield neural networks (HNNs) [5] define an energy function derived from an objective function, and the state of the

HNN changes in accordance with the energy function until the state becomes stable. A solution is then derived from the stable state. Boltzmann machines [6] are variations of HNNs to which a stochastic parameter is added so that the state first changes stochastically, and the stochastic changes in the state gradually decrease. With this parameter, the state falls into a local optimum less frequently. Simulated annealing (SA) [7] is a random search method that was developed for performing random searches more effectively. It uses the principle of annealing from metal engineering: slowly cooling heated metal produces a superior crystalline structure. Genetic algorithms (GAs) [8] are also random search methods; they use the principle of organic evolution to achieve the same thing. A solution is considered to be an individual living creature, and it is improved by using mutation and combinations of information for various individuals. Although these methods have been well developed, they unfortunately cannot always derive the global optimum.

These methods are designed to avoid being trapped in a local optimum by using random variables, chaos, the structure of the objective function, or the relationship between local optimums. However, there is a fundamental limit to their ability because they search for the global optimum in a real space. Quantum mechanics was recently added to some optimization methods to overcome this difficulty. Instead of SA thermal fluctuations in a real space, quantum annealing (QA) uses quantum fluctuations and thus has a shorter convergence time [9], [10]. Quantum neural networks (QNNs) [11], [12] are variations of HNNs and were developed to effectively perform a full search on the basis of the superposition of quantum states. Because QA and QNNs use the properties of a quantum mechanical system, they can find the global optimum without falling into a local optimum. However, it is very difficult to observe the quantum dynamics of a wave function and to construct a device to create the quantum effects that are required for QA and QNNs. It is thus still difficult to apply them to most optimization problems.

A method for global optimization has been developed that overcomes the defects of the methods described above. A nonlinear objective function with multiple local optimums is expanded into a linear function of a wave coefficient vector, and the problem of maximizing the objective function is reduced to that of

Manuscript received June 11, 2009.

Manuscript revised August 19, 2009.

[†]The author is with Future University Hakodate, Hakodate-shi, 041-8655 Japan.

maximizing a quadratic objective function. Because it uses the wave function [13] expressed by the wave coefficients, it is equivalent to a full search method, i.e., one that searches in parallel for the global optimum in the whole domain of definition. Therefore, it can always find the approximate global optimum. The method is presented in this paper, and its validity is demonstrated using various objective functions. Computer simulation showed that it can always produce a good approximation of the global optimum for each objective function if there is a unique global optimum.

2. Wave Coefficient Equation for Nonlinear Objective Function

A wave coefficient equation (WCE) was developed to reduce the problem of maximizing a nonlinear objective function to that of maximizing a quadratic function with respect to the wave coefficients. The moment vector equation (MVE) [14] of the objective function is first derived and is then transformed into the WCE. The MVE is summarized in Sect. 2.1, and the WCE for the objective function is presented in Sect. 2.2.

2.1 Moment Vector Equation

The MVE was developed to approximate an arbitrary multi-dimensional nonlinear function in the whole domain of definition [14]. Consider the following nonlinear function:

$$y = f(\mathbf{x}), \quad (1)$$

where $\mathbf{x} \stackrel{\text{def}}{=} (x_1, \dots, x_{d_x})^T \in \mathcal{D}_{\mathbf{x}}$ is the state vector of dimension d_x , $\mathcal{D}_{\mathbf{x}} \stackrel{\text{def}}{=} \{\mathbf{x} | x_{\min d} < x_d < x_{\max d}, 1 \leq d \leq d_x\}$ is the domain of the definition of \mathbf{x} , $y \in \mathcal{D}_{\mathbf{y}}$ is the value of $f(\mathbf{x})$, $\mathcal{D}_{\mathbf{y}} \stackrel{\text{def}}{=} \{y | y_{\min} < y < y_{\max}\}$ is the domain of the definition of y , and superscript T denotes transposition. If the domain of definitions cannot be set in advance, we can modify Eq. (1) to obtain

$$y = h_y(f(\mathbf{h}_x(\mathbf{x}))), \quad (2)$$

where $h_y(y')$ is a monotone increasing function such that $y_{\min} \leq h_y(y') \leq y_{\max}$ for $-\infty \leq y' \leq \infty$, $\mathbf{h}_x(\mathbf{x}) \stackrel{\text{def}}{=} (h_{x1}(x_1), \dots, h_{xd_x}(x_{d_x}))^T$, and $h_{xd}(x_d)$ is a monotone increasing function such that $-\infty \leq h_{xd}(x_d) \leq \infty$ for $x_{\min d} \leq x_d \leq x_{\max d}$.

Let $\{\psi_i(y)\}$ and $\{\psi_i(\mathbf{x})\}$ be orthonormal bases defined in Appendix A. Note that the same symbol, ψ , is used to simplify the explanation, although $\{\psi_i(y)\}$ and $\{\psi_i(\mathbf{x})\}$ are generally different bases. To derive the MVE for the nonlinear function in Eq. (1), the following assumption is introduced with respect to Eq. (1).

Assumption 1: We can expand $E[\psi_i(y)|\mathbf{x}]$ into a Fourier series:

$$E[\psi_i(y)|\mathbf{x}] = \sum_{j=0}^{N_x} a_{ij} \psi_j(\mathbf{x}) + \varepsilon_i(\mathbf{x}), \quad (3)$$

where N_x is the degree of expansion of $E[\psi_i(y)]$, $E[\cdot]$ is the mathematical expectation, and $\varepsilon_i(\mathbf{x})$ is the residual. \square

Using Eq. (3), we can expand $E[\psi_i(y)]$:

$$\begin{aligned} E[\psi_i(y)] &= \int \psi'_i p(\psi'_i) d\psi'_i \\ &= \int \psi'_i \int p(\mathbf{x}) p(\psi'_i|\mathbf{x}) d\mathbf{x} d\psi'_i \\ &= \int p(\mathbf{x}) \int \psi'_i p(\psi'_i|\mathbf{x}) d\psi'_i d\mathbf{x} \\ &= \int p(\mathbf{x}) E[\psi'_i|\mathbf{x}] d\mathbf{x} \\ &= \int p(\mathbf{x}) \left(\sum_{j=0}^{N_x} a_{ij} \psi_j(\mathbf{x}) + \varepsilon_i(\mathbf{x}) \right) d\mathbf{x} \\ &= \sum_{j=0}^{N_x} a_{ij} E[\psi_j(\mathbf{x})] + E[\varepsilon_i(\mathbf{x})], \end{aligned} \quad (4)$$

where ψ'_i denotes $\psi_i(y)$ and $p(\mathbf{x})$ denotes the probability density function (pdf) of \mathbf{x} . When Eq. (1) is deterministic, a_{ij} is obtained using Eq. (A.2):

$$a_{ij} = \int_{\mathcal{D}_{\mathbf{x}}} \psi_i(f(\mathbf{x})) \psi_j^*(\mathbf{x}) d\mathbf{x}, \quad (5)$$

where superscript * denotes a complex conjugate. If we assume that $E[\varepsilon_i(\mathbf{x})] = 0$, Eq. (4) can be expressed using a linear function:

$$E[\psi(y)] = A_{\psi} E[\psi(\mathbf{x})]. \quad (6)$$

This equation is referred to as the MVE, $\psi(y) \stackrel{\text{def}}{=} (\psi_0(y), \dots, \psi_{N_y}(y))^T$, N_y is the degree of expansion of y , $\psi(\mathbf{x}) \stackrel{\text{def}}{=} (\psi_0(\mathbf{x}), \dots, \psi_{N_x}(\mathbf{x}))^T$, and

$$A_{\psi} \stackrel{\text{def}}{=} \begin{bmatrix} a_{00} & \cdots & a_{0N_x} \\ \vdots & \ddots & \vdots \\ a_{N_y 0} & \cdots & a_{N_y N_x} \end{bmatrix}.$$

The nonlinear function in Eq. (1) is approximately expressed by the MVE in Eq. (6). The accuracy of Eq. (6) increases as N_x and N_y increase. Using Eq. (6), we can derive not only the expected value of $\psi_i(y)$ but also the statistical properties such as the mean, variance, covariance, and pdf of y [14].

2.2 Wave Coefficient Equation

Let $\Psi(\mathbf{x})$ be a wave function [13]. We can expand $\Psi(\mathbf{x})$ using orthonormal basis $\{\psi_i(\mathbf{x})\}$:

$$\begin{aligned}\Psi(\mathbf{x}) &\cong \sum_{i=0}^{N_x} c_i \psi_i(\mathbf{x}) \\ &= \mathbf{c}^T \boldsymbol{\psi}(\mathbf{x}),\end{aligned}\quad (7)$$

where c_i is the expansion coefficient of the wave function, which is referred to as the wave coefficient in this paper, and $\mathbf{c} \stackrel{\text{def}}{=} (c_0, \dots, c_{N_x})^T$. As shown in Appendix A, the wave coefficient can be obtained using

$$c_i = \int_{\mathcal{D}_x} \Psi(\mathbf{x}) \psi_i^*(\mathbf{x}) d\mathbf{x}. \quad (8)$$

Probability density function $p(\mathbf{x})$ is obtained using $\Psi(\mathbf{x})$ [13]:

$$\begin{aligned}p(\mathbf{x}) &= \Psi(\mathbf{x}) \Psi^*(\mathbf{x}) \\ &\cong \boldsymbol{\psi}^T(\mathbf{x}) \mathbf{c} \mathbf{c}^\dagger \boldsymbol{\psi}^*(\mathbf{x}),\end{aligned}\quad (9)$$

where superscript \dagger denotes conjugate transposition. Because $\int p(\mathbf{x}) d\mathbf{x} = \int \Psi(\mathbf{x}) \Psi^*(\mathbf{x}) d\mathbf{x} = 1$ and basis $\{\psi_i(\mathbf{x})\}$ is orthonormal, the following equation holds [13]:

$$\|\mathbf{c}\| = 1,$$

where $\|\mathbf{c}\| \stackrel{\text{def}}{=} \sqrt{\mathbf{c}^T \mathbf{c}^*}$.

Consider the pdfs defined by $p(\mathbf{x}) \stackrel{\text{def}}{=} \delta(\mathbf{x} - \hat{\mathbf{x}})$ and $p(y) \stackrel{\text{def}}{=} \delta(y - \hat{y})$. Let $\hat{\mathbf{q}}$ be the wave coefficient vector of $p(\mathbf{x})$ and $\hat{\mathbf{r}}$ be that of $p(y)$. Using Eq. (A.9) in Appendix B, we can obtain the wave coefficient vectors by using

$$\hat{\mathbf{q}} \stackrel{\text{def}}{=} \xi_q^{-1} \boldsymbol{\psi}^*(\hat{\mathbf{x}}), \quad (10)$$

$$\hat{\mathbf{r}} \stackrel{\text{def}}{=} \xi_r^{-1} \boldsymbol{\psi}^*(\hat{y}), \quad (11)$$

where ξ_q and ξ_r are defined in the same manner as in Eq. (A.10). The normalization condition with respect to the wave coefficient is hereinafter eliminated. That is, $\|\hat{\mathbf{q}}\| = 1$ and $\|\hat{\mathbf{r}}\| = 1$ do not always hold[†]. From Eq. (9), we obtain

$$p(\mathbf{x}) \cong \|\hat{\mathbf{q}}\|^{-2} \boldsymbol{\psi}^T(\mathbf{x}) \hat{\mathbf{q}} \hat{\mathbf{q}}^\dagger \boldsymbol{\psi}^*(\mathbf{x}), \quad (12)$$

$$p(y) \cong \|\hat{\mathbf{r}}\|^{-2} \boldsymbol{\psi}^T(y) \hat{\mathbf{r}} \hat{\mathbf{r}}^\dagger \boldsymbol{\psi}^*(y). \quad (13)$$

We can modify Eq. (6) to obtain

$$\xi_r^{-1} E[\boldsymbol{\psi}^*(y)] = \xi_r^{-1} \xi_q A_\psi^* \xi_q^{-1} E[\boldsymbol{\psi}^*(\mathbf{x})].$$

Because $p(\mathbf{x}) = \delta(\mathbf{x} - \hat{\mathbf{x}})$ and $p(y) = \delta(y - \hat{y})$, $\boldsymbol{\psi}^*(\hat{\mathbf{x}}) = E[\boldsymbol{\psi}^*(\mathbf{x})]$ and $\boldsymbol{\psi}^*(\hat{y}) = E[\boldsymbol{\psi}^*(y)]$. Thus, by substituting $A \stackrel{\text{def}}{=} \xi_r^{-1} \xi_q A_\psi^*$ and Eqs. (10) and (11) into the above

[†]Even if $\|\hat{\mathbf{q}}\| = 1$, $\|\hat{\mathbf{r}}\| = 1$ does not always hold in Eq. (14) because matrix A is not a unitary matrix. Thus, the normalization conditions with respect to $\hat{\mathbf{q}}$ and $\hat{\mathbf{r}}$ were omitted.

equation, we obtain

$$\hat{\mathbf{r}} = A \hat{\mathbf{q}}. \quad (14)$$

This equation is an approximation of Eq. (1) based on the (unnormalized) wave coefficient vectors and yields $p(y)$ from $p(\mathbf{x})$ assuming that $p(\mathbf{x})$ and $p(y)$ are delta functions.

Let us generalize Eq. (14) for an arbitrary pdf. Consider the following equation for arbitrary wave coefficient vectors \mathbf{q} and \mathbf{r} :

$$\mathbf{r} = A \mathbf{q}. \quad (15)$$

This equation is referred to as the wave coefficient equation (WCE) in this paper. As shown in Appendix C, $p(\mathbf{x})$ is obtained from \mathbf{q} , \mathbf{r} is obtained from \mathbf{q} as in Eq. (15), $p(y)$, which is related to $p(\mathbf{x})$ by Eq. (1), is obtained from \mathbf{r} , and matrix A in Eq. (15) is obtained from Eq. (1). Thus, Eq. (14) can be generalized as Eq. (15), which expresses the nonlinear function in Eq. (1) for arbitrary \mathbf{q} , and $p(\mathbf{x})$ and $p(y)$ are obtained using

$$p(\mathbf{x}) \cong \|\mathbf{q}\|^{-2} \boldsymbol{\psi}^T(\mathbf{x}) \mathbf{q} \mathbf{q}^\dagger \boldsymbol{\psi}^*(\mathbf{x}), \quad (16)$$

$$p(y) \cong \|\mathbf{r}\|^{-2} \boldsymbol{\psi}^T(y) \mathbf{r} \mathbf{r}^\dagger \boldsymbol{\psi}^*(y). \quad (17)$$

3. Global Optimization in Wave Coefficient Space

A global optimization method based on the WCE in Eq. (15) is presented in this section.

Consider the optimization problem of maximizing $y \in \mathcal{D}_y$ obtained by Eq. (1) for $\mathbf{x} \in \mathcal{D}_x$. Using WCE in Eq. (15) and pdf in Eq. (17), we can rewrite the optimization problem as the problem of maximizing $E[y]$ defined by

$$\begin{aligned}E[y] &\stackrel{\text{def}}{=} \int y p(y) dy \\ &= \int y \|\mathbf{r}\|^{-2} \boldsymbol{\psi}^T(y) \mathbf{r} \mathbf{r}^\dagger \boldsymbol{\psi}^*(y) dy \\ &= \|A \mathbf{q}\|^{-2} \mathbf{q}^T A^T Y A^* \mathbf{q}^*,\end{aligned}\quad (18)$$

where $Y \stackrel{\text{def}}{=} \int y \boldsymbol{\psi}(y) \boldsymbol{\psi}^\dagger(y) dy$ is an $(N_y + 1) \times (N_y + 1)$ matrix (See Appendix D for details). If $\|A \mathbf{q}\|^2 = \mathbf{q}^T A^T A^* \mathbf{q}^* = 1$, Eq. (18) reduces to $\mathbf{q}^T A^T Y A^* \mathbf{q}^*$. Thus, the optimization problem of maximizing $E[y]$ is expressed as

$$\text{Object : } \max_{\mathbf{q}} \mathbf{q}^T A^T Y A^* \mathbf{q}^*, \quad (19)$$

$$\text{Constraint : } \mathbf{q}^T A^T A^* \mathbf{q}^* = 1. \quad (20)$$

Because the objective function of this problem is a quadratic equation with respect to \mathbf{q} , we can solve

the above optimization problem using the steepest descent method (SDM) in a wave coefficient space for the following Lagrange function [17]:

$$L(\mathbf{q}, \mu) \stackrel{\text{def}}{=} -\mathbf{q}^T A^T Y A^* \mathbf{q}^* + \mu(\mathbf{q}^T A^T A^* \mathbf{q}^* - 1), \quad (21)$$

where $\mu(-\infty \leq \mu \leq \infty)$ is the Lagrange multiplier. The algorithm is described in Algorithm 1.

Algorithm 1: Steepest descent method in a wave coefficient space for Lagrange function $L(\mathbf{q}, \mu)$.

- (1-1) Set $t = 0$.
- (1-2) Set step sizes α_q , α_μ , and α_{step} (all > 0) and initial values \mathbf{q}_0 and μ_0 .
- (1-3) Compute \mathbf{d}_{qt} and $\mathbf{d}_{\mu t}$:
 $\mathbf{d}_{qt} = -\nabla_{\text{Re}[\mathbf{q}]} L(\mathbf{q}_t, \mu_t) - i \nabla_{\text{Im}[\mathbf{q}]} L(\mathbf{q}_t, \mu_t)$,
 $\mathbf{d}_{\mu t} = -\nabla_{\mu} L(\mathbf{q}_t, \mu_t)$.
- (1-4) Compute \mathbf{q}_{t+1} and μ_{t+1} :
 $\mathbf{q}_{t+1} = \mathbf{q}_t + \alpha_q \mathbf{d}_{qt}$,
 $\mu_{t+1} = \mu_t + \alpha_\mu \mathbf{d}_{\mu t}$.
- (1-5) If $R(t, \alpha_{\text{step}}) = 0$ and $J_{\text{opt}}(t) - J_{\text{opt}}(t - \alpha_{\text{step}}) < 0$, set $\tilde{\mathbf{q}}_{\text{opt}} = \mathbf{q}_t$ and finish the algorithm.
- (1-6) Set $t = t + 1$ and go to Step (1-2).

Here, $R(t, \alpha_{\text{step}})$ is the remainder of t upon division by α_{step} , $J_{\text{opt}}(t) \stackrel{\text{def}}{=} E[y]_t / (\sum_d \sigma[x_d]_t)$, α_{step} is set to a natural number to eliminate small fluctuations in $J_{\text{opt}}(t)$, $\nabla_{\mathbf{x}}$ denotes the nabla operator with respect to \mathbf{x} , $E[y]_t$ is defined using Eq. (18) as

$$E[y]_t \stackrel{\text{def}}{=} \|A\mathbf{q}_t\|^{-2} \mathbf{q}_t^T A^T Y A^* \mathbf{q}_t^*, \quad (22)$$

and $\sigma[x_d]_t$ is defined as

$$\sigma[x_d]_t \stackrel{\text{def}}{=} \sqrt{E[x_d^2]_t - E[x_d^1]_t^2}, \quad (23)$$

$$\begin{aligned} E[x_d^n]_t &= \int x_d^n p(\mathbf{x}) d\mathbf{x} \\ &\cong \int x_d^n \|\mathbf{q}_t\|^{-2} \mathbf{q}_t^T \boldsymbol{\psi}(\mathbf{x}) \boldsymbol{\psi}^\dagger(\mathbf{x}) \mathbf{q}_t^* d\mathbf{x} \\ &= \|\mathbf{q}_t\|^{-2} \mathbf{q}_t^T X_{d(n)} \mathbf{q}_t^*, \end{aligned} \quad (24)$$

$$X_{d(n)} \stackrel{\text{def}}{=} \int x_d^n \boldsymbol{\psi}(\mathbf{x}) \boldsymbol{\psi}^\dagger(\mathbf{x}) d\mathbf{x}. \quad (25)$$

The reason the condition for finishing the algorithm in Step (1-5) is used will be discussed in Sect. 4.2.

The approximation of the global optimum, $E[x_d]_{\text{opt}}$, which approximately maximizes Eq. (1), and the approximate maximum of Eq. (1), $E[y]_{\text{opt}}$, are expressed using $\tilde{\mathbf{q}}_{\text{opt}}$ obtained using Algorithm 1:

$$E[x_d]_{\text{opt}} = \|\tilde{\mathbf{q}}_{\text{opt}}\|^{-2} \tilde{\mathbf{q}}_{\text{opt}}^T X_{d(1)} \tilde{\mathbf{q}}_{\text{opt}}^*, \quad (26)$$

$$E[y]_{\text{opt}} \stackrel{\text{def}}{=} \|A\tilde{\mathbf{q}}_{\text{opt}}\|^{-2} \tilde{\mathbf{q}}_{\text{opt}}^T A^T Y A^* \tilde{\mathbf{q}}_{\text{opt}}^*. \quad (27)$$

The calculation cost of Algorithm 1 is not lower than that of conventional methods because Algorithm 1 is based on the steepest decent method. Although we can reduce its calculation cost using the property of the objective function in Eq. (19), which is a convex quadratic function, it is very difficult to obtain higher performance with respect to the calculation cost compared with conventional methods, which have been improved over a long period of time through various applications. Two properties of Algorithm 1, the global optimum is always obtained (as shown in Sect. 4) and Algorithm 1 uses a wave function, are more important than the calculation cost because they are also properties of quantum computing [15].

That is, the solution described in Eqs. (26) and (27) is obtained using the wave coefficient vector, which describes a quantum state. This is the same as in quantum computing. However, there are differences between the method presented in this paper and quantum computing.

- Quantum mechanics describes a linear relationship in quantum states.
- Quantum computing transforms input quantum state \mathbf{c} to output quantum state \mathbf{c}' using unitary operator U . Thus, $\|\mathbf{c}\| = \|\mathbf{c}'\|$ always holds.
- In contrast, the WCE in Eq. (15) describes a nonlinear objective function in Eq. (1), so matrix A is not unitary. That is, $\|\mathbf{r}\| = \|\mathbf{q}\|$ does not always hold.
- This means that the WCE in Eq. (15) expresses a nonlinear relationship in real space by omitting the restriction of unitary representations from matrix A .

It is thus difficult to apply the WCE to quantum computing directly. The application of the method presented in this paper to quantum computing is left for future work.

4. Performance Evaluation

Algorithm 1 presented in this paper reduces the optimization problem for a nonlinear objective function in Eq. (1) to that for a convex quadratic objective function with a convex quadratic equality constraint in Eqs. (19) and (20). Generally, we cannot always derive a global optimum for both functions. However, Algorithm 1 uses a wave function expressed in a wave coefficient space and a distribution of global optima for Eq. (1) can be expressed by a pdf with a wave coefficient vector regardless of the number of global optima. Algorithm 1 is thus equivalent to a full search algorithm if its initial state expresses a uniform distribution. Therefore, a global optimum should always be obtained. In this section, Algorithm 1 is examined for various objective functions to evaluate its performance.

Table 1 Parameters of $f_G(x_1)$ used to examine condition for finishing Algorithm 1.

ℓ	1	2	3	4	5
β_ℓ	0.50	0.85	0.50	0.70	0.60
$\gamma_{1\ell}$	0.08	0.20	0.35	0.55	0.80
$\zeta_{1\ell}$	0.05	0.05	0.05	0.15	0.10

4.1 Objective Functions

Consider the problems of maximizing Gaussian-type function $f_G(\mathbf{x})$ and square-type function $f_S(\mathbf{x})$, defined by

$$f_G(\mathbf{x}) \stackrel{\text{def}}{=} \alpha + \sum_{\ell=1}^{N_{\text{extrm}}} \beta_\ell \prod_{d=1}^{d_x} \exp\left(-\frac{(x_d - \gamma_{d\ell})^2}{\zeta_{d\ell}^2}\right), \quad (28)$$

$$f_S(\mathbf{x}) \stackrel{\text{def}}{=} \alpha + \sum_{\ell=1}^{N_{\text{extrm}}} \beta_\ell \prod_{d=1}^{d_x} \text{squ}(x_d, \gamma_{d\ell}, \zeta_{d\ell}), \quad (29)$$

where each $f_G(\mathbf{x})$ and $f_S(\mathbf{x})$ is the superposition of functions with a unique extreme, N_{extrm} is the number of the superposition, α is the minimum value of $f_G(\mathbf{x})$ and $f_S(\mathbf{x})$, β_ℓ is the weight of the ℓ th extreme value, $\gamma_{d\ell}$ is the coordinate of the ℓ th extreme value on the x_d -axis, $\zeta_{d\ell}$ is the width of the ℓ th extreme value, and $\text{squ}(x_d, \gamma_{d\ell}, \zeta_{d\ell})$ is a one-dimensional square function defined by

$$\text{squ}(x_d, \gamma_{d\ell}, \zeta_{d\ell}) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \gamma_{d\ell} - \frac{\zeta_{d\ell}}{2} < x_d \leq \gamma_{d\ell} + \frac{\zeta_{d\ell}}{2} \\ 0 & \text{otherwise.} \end{cases}$$

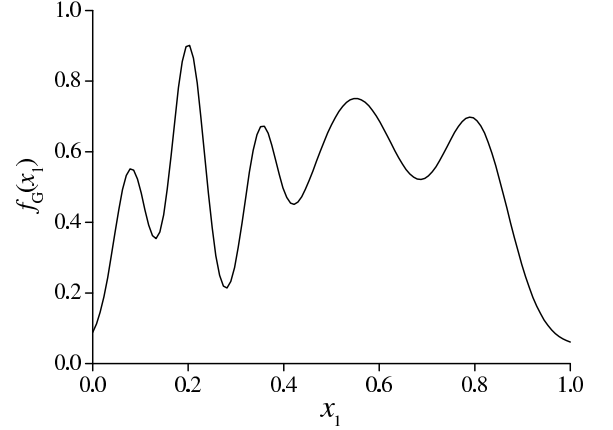
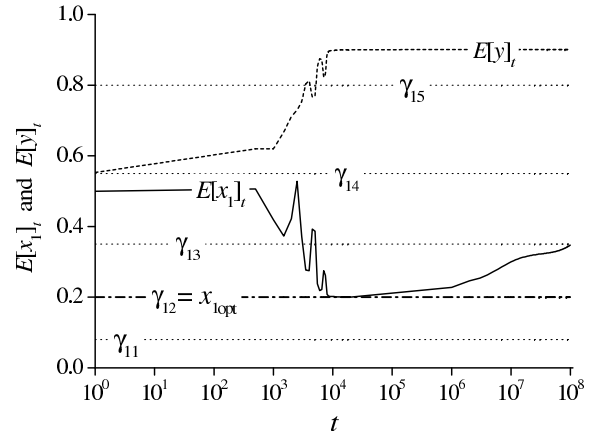
4.2 Condition for Finishing Algorithm 1

Consider one-dimensional Gaussian-type function $f_G(x_1)$ with $N_{\text{extrm}} = 5$, $d_x = 1$, $\alpha = 0.05$, $0 \leq x_1 \leq 1.0$, $0 \leq y \leq 1.0$, and the parameters in Table 1. As we can see from Table 1 and the objective function shown in Fig. 1, there are four local optimums, and the global optimum, $x_{1\text{opt}}$, is equal to 0.2 ($= \gamma_{12}$). Using Algorithm 1 with $N_x = N_y = 128$, we can derive the global optimum. Here, $\alpha_q = 10^{-3}$, $\alpha_\mu = 10^{-3}$, and $\alpha_{\text{step}} = 10^4$. Figure 2 shows the changes in $E[y]_t$ and $E[x_1]_t$, which are defined by Eqs. (22) and (24), respectively. Figures 3 and 4 show the changes in $p_t(\mathbf{x})$ and $p_t(\mathbf{y})$, respectively, which are the pdfs at the t th step defined by

$$p_t(\mathbf{x}) \stackrel{\text{def}}{=} \|\mathbf{q}_t\|^{-2} \psi^T(\mathbf{x}) \mathbf{q}_t \mathbf{q}_t^\dagger \psi^*(\mathbf{x}), \quad (30)$$

$$p_t(\mathbf{y}) \stackrel{\text{def}}{=} \|A\mathbf{q}_t\|^{-2} \psi^T(\mathbf{x}) A\mathbf{q}_t \mathbf{q}_t^\dagger A^\dagger \psi^*(\mathbf{x}). \quad (31)$$

The \mathbf{q}_0 and μ_0 values are initialized: $\mathbf{q}_0 = (1, 0, \dots, 0)^T$ and $\mu_0 = 0$. The former means that $p_0(x_1)$ is a uniform distribution. Here, after the condition for finishing the algorithm is satisfied, the algorithm is continued until $t = 10^8$ to investigate the

**Fig. 1** $f_G(x_1)$ defined in Table 1**Fig. 2** Changes in $E[x_1]_t$ and $E[y]_t$ obtained using Algorithm 1 for $f_G(x_1)$ in Fig. 1.

convergence process.

As shown in Fig. 2, $E[x_1]_t$ changes from $E[x_1]_0 = 0.5$, which is the mean value of the uniform distribution, toward the global optimum ($x_{1\text{opt}} = \gamma_{12}$) passing through the local optimum (γ_{13}), and reaches the global optimum at about $t = 10^4$. We can see in Fig. 3 the change in $p_t(x_1)$; Algorithm 1 takes the whole value of x_1 into account in accordance with $p_t(x_1)$ at each step rather than investigates $f_G(x_1)$ for various values of x_1 step by step.

Once $E[x_1]_t$ converges to global optimum $x_{1\text{opt}}$ at about $t = 10^4$, it separates from $x_{1\text{opt}}$, as shown in Fig. 2. Because Eqs. (19) and (20) are quadratic forms, it is clear that there are many values of \mathbf{q} that satisfy Eqs. (19) and (20). However all values of \mathbf{q} do not necessarily satisfy $\|\mathbf{q}\| = 1$ because matrix A is not a unitary matrix. Algorithm 1 thus can select various values of \mathbf{q} as a solution even if they do not satisfy $\|\mathbf{q}\| = 1$. This means that the solution may not express the correct pdf of x_1 . This behavior of Algorithm 1 can be confirmed by comparing Figs. 2 through 5. The shape of $p_t(x_1)$ once converges to a delta function, which is the

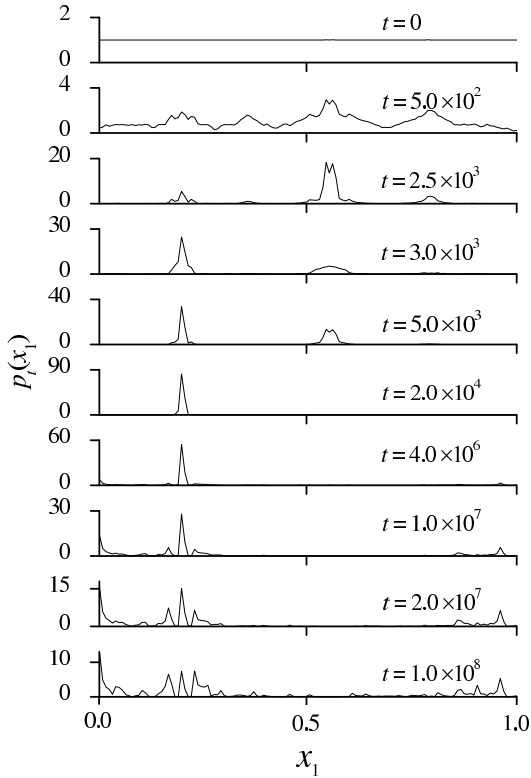


Fig. 3 Change in $p_t(x_1)$ obtained using Algorithm 1 for $f_G(x_1)$ in Fig. 1.

correct pdf of x_1 when the optimal value is obtained, at about $t = 10^4$, as shown in Fig. 3. As shown in Fig. 5, $\|\mathbf{q}_t\|$ and $\|\mathbf{Aq}_t\|$ once converge to almost 1 at the same time. However, $\|\mathbf{q}_t\|$ separates from 1 whereas $\|\mathbf{Aq}_t\|$ remains near 1. The reason for this latter behavior is that Eq. (20) works as a constraint in Algorithm 1, and the reason for the former behavior is that matrix A is not a unitary matrix. The value of $E[x_1]_t$ separates from $x_{1\text{opt}}$, as shown in Fig. 2, and the shape of $p_t(x_1)$ collapses as $\|\mathbf{q}_t\|$ separates from 1, as shown in Fig. 3. However, $E[y]_t$ takes the maximum value even though $\|\mathbf{q}_t\|$ separates from 1, as shown in Fig. 2. It is thus clear that these changes in $E[x_1]_t$ and $p_t(x_1)$ arise from the change in $\|\mathbf{q}_t\|$ and that there are many values of \mathbf{q}_t that maximize $E[y]_t$ even though it does not express an appropriate pdf. Figure 4 shows that $p_t(y)$ retains the delta function while $E[y]_t$ takes a maximum value even though $p_t(x_1)$ is separate from the delta function. This means that the optimization problem for y can be replaced with that for $E[y]$, as described in Appendix D.

As mentioned above, $E[x_1]_t$ obtained by Algorithm 1 does not always express the optimal value even if $E[y]_t$ takes the maximum value. We can find whether \mathbf{q}_t obtained by Algorithm 1 expresses an appropriate pdf by using the sum of the differences between the norm and 1, defined by $|\|\mathbf{q}_t\| - 1| + |\|\mathbf{Aq}_t\| - 1|$. It becomes almost zero, as shown in Fig. 5, if the pdf is appropriate.

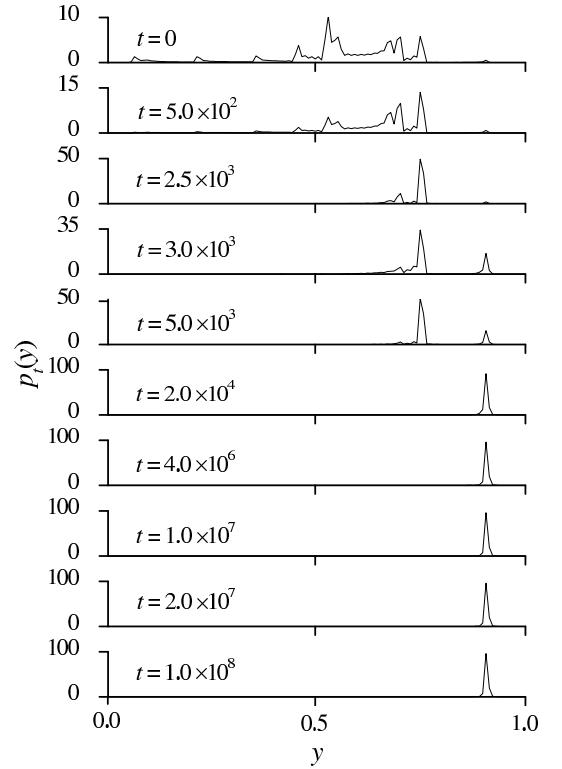


Fig. 4 Change in $p_t(y)$ obtained using Algorithm 1 for $f_G(x_1)$ in Fig. 1.

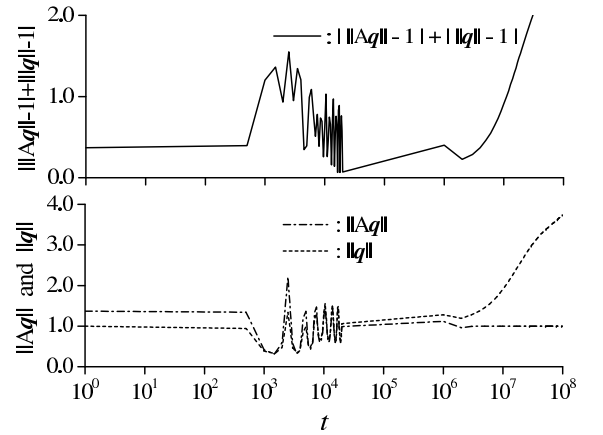


Fig. 5 Change in norm of wave coefficient vector obtained using Algorithm 1 for $f_G(x_1)$ in Fig. 1.

However, it cannot be used as the condition for deciding whether $E[x_1]_t$ is optimum because $E[x_1]_t$ is not always optimum when it is equal to zero.

To investigate the condition for finishing the algorithm immediately after $E[x_1]_t$ converges to $x_{1\text{opt}}$, the changes in $E[y]_t$, $\sigma[x_1]_t$, and $E[y]_t/\sigma[x_1]_t$ were evaluated, as shown in Fig. 6. Algorithm 1 maximizes $E[y]_t$, and $E[y]_t$ remains maximum after $E[x_1]_t$ is separated from the global optimum. In this sense, Algorithm 1 works well. However, we can see in Figs. 3 and 6 that various shapes of $p_t(\mathbf{x})$ yield the maximum value. That

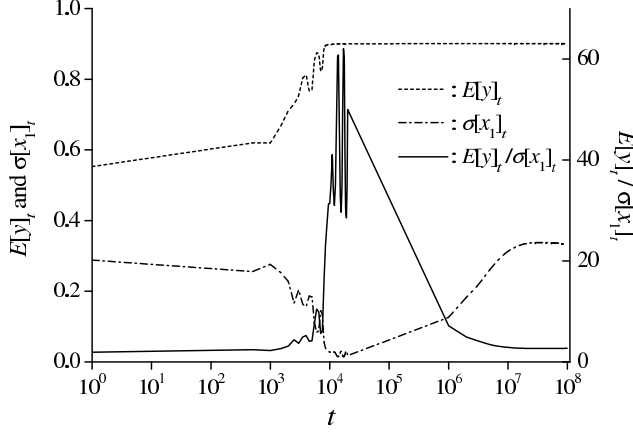


Fig. 6 Changes in $E[y]_t$, $\sigma[x_1]_t$, and $E[y]_t/\sigma[x_1]_t$ obtained using Algorithm 1 for $f_G(x_1)$ in Fig. 1.

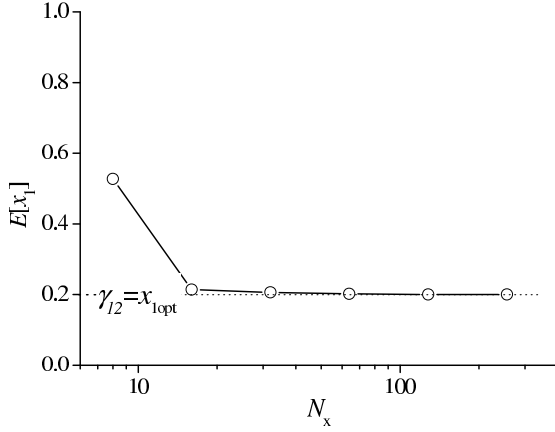


Fig. 7 Effect of N_x on accuracy of solution for $f_G(x_1)$ in Fig. 1.

is, there are many solutions of \mathbf{q} that maximize $E[y]_t$. Therefore, the value of $E[y]_t$ should not be used as the condition for finishing Algorithm 1. In contrast, $E[y]_t/\sigma[x_1]_t$ is almost maximum and $\sigma[x_1]_t$ is almost minimum when $E[x_1]_t \cong x_{1\text{opt}}$. Because various evaluations (the results are omitted in this paper) showed that $E[y]_t/\sigma[x_1]_t$ is superior to $\sigma[x_1]_t$ from the viewpoint of solution accuracy, $E[y]_t/\sigma[x_1]_t$ is used as the condition for finishing Algorithm 1 (Step (1-5) in Algorithm 1).

4.3 Evaluation

The effect of N_x on the accuracy of the solution was evaluated for one-dimensional Gaussian-type function $f_G(\mathbf{x})$, which is the same function used in Sect. 4.2. Here, N_y is set equal to N_x . As shown in Fig. 7, good approximations of the global optimum were obtained when $N_x \geq 16$, and the accuracy of the approximations increased with the value of N_x .

Consider two-dimensional Gaussian-type function

Table 2 Corresponding table for functions, parameter table, and figures of functions and $\tilde{p}_{\text{opt}}(\mathbf{x})$.

Function	Parameter table	Function	$\tilde{p}_{\text{opt}}(\mathbf{x})$
$f_G(\mathbf{x}) _{\text{uniGO}}$	Table 3	Fig. 8	Fig. 9
$f_S(\mathbf{x}) _{\text{uniRGO}}$	Table 3	Fig. 10	Fig. 11
$f_G(\mathbf{x}) _{\text{twoGO}}$	Table 4	Fig. 12	Fig. 13

Table 3 Parameters for $f_G(\mathbf{x})$ with unique global optimum and $f_S(\mathbf{x})$ with unique region of global optimum.

ℓ	1	2	3	4
β_ℓ	0.60	0.85	0.40	0.20
$\gamma_{1\ell}$	0.20	0.80	0.20	0.80
$\zeta_{1\ell}$	0.20	0.20	0.20	0.20
$\gamma_{2\ell}$	0.20	0.20	0.80	0.80
$\zeta_{2\ell}$	0.20	0.20	0.20	0.20

Table 4 Parameters for $f_G(\mathbf{x})$ with two global optimums.

ℓ	1	2	3	4
β_ℓ	0.60	0.85	0.85	0.20
$\gamma_{1\ell}$	0.20	0.80	0.20	0.80
$\zeta_{1\ell}$	0.20	0.20	0.20	0.20
$\gamma_{2\ell}$	0.20	0.20	0.80	0.80
$\zeta_{2\ell}$	0.20	0.20	0.20	0.20

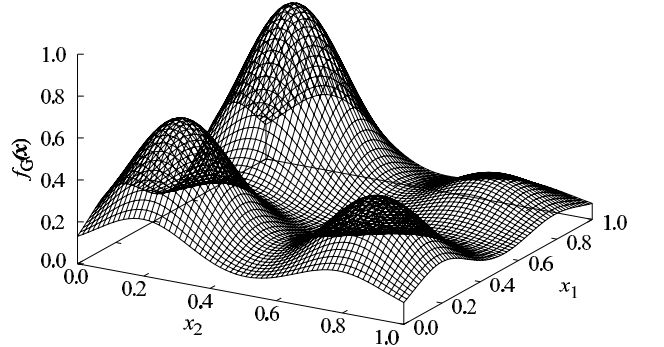


Fig. 8 $f_G(\mathbf{x})|_{\text{uniGO}}$ defined in Table 2.

$f_G(\mathbf{x})$ and two-dimensional square-type function $f_S(\mathbf{x})$ with $N_{\text{extrm}} = 4$, $d_x = 2$, $\alpha = 0.05$, $0 \leq x_1 \leq 1.0$, $0 \leq x_2 \leq 1.0$, $0 \leq y \leq 1.0$, and the parameters in Table 3 or 4. The relationship between the function, the parameter table, the figure of the function, and the figure of $\tilde{p}_{\text{opt}}(\mathbf{x})$ defined by

$$\tilde{p}_{\text{opt}}(\mathbf{x}) \stackrel{\text{def}}{=} \|\tilde{\mathbf{q}}_{\text{opt}}\|^{-2} \psi^T(\mathbf{x}) \tilde{\mathbf{q}}_{\text{opt}} \tilde{\mathbf{q}}_{\text{opt}}^\dagger \psi^*(\mathbf{x}) \quad (32)$$

is shown in Table 2, where $f_G(\mathbf{x})|_{\text{uniGO}}$ denotes $f_G(\mathbf{x})$ with a unique global optimum, $f_S(\mathbf{x})|_{\text{uniRGO}}$ denotes $f_S(\mathbf{x})$ with a unique region in which $\forall \mathbf{x}$ are global optimums, and $f_G(\mathbf{x})|_{\text{twoGO}}$ denotes $f_G(\mathbf{x})$ with two global optimums.

Table 5 shows the approximations of the global optimums for the three functions in Table 2. As shown in Table 5, the approximation of the global optimum for $f_G(\mathbf{x})|_{\text{uniGO}}$ obtained using Algorithm 1 ($E[\mathbf{x}]_{\text{opt}} \stackrel{\text{def}}{=} (E[x_1]_{\text{opt}}, E[x_2]_{\text{opt}})^T$) and Eq. (26) is close

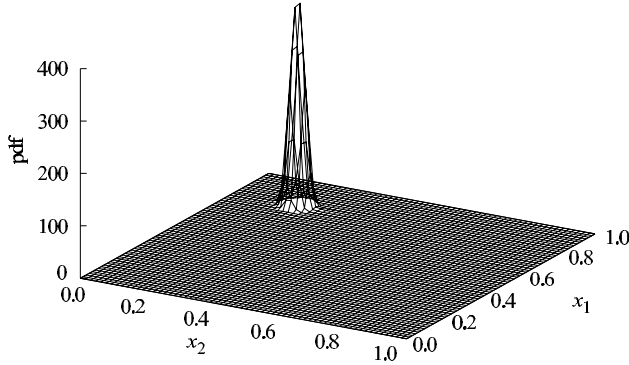


Fig. 9 $\tilde{p}_{\text{opt}}(\mathbf{x})$ for $f_G(\mathbf{x})|_{\text{uniGO}}$ obtained using Algorithm 1.

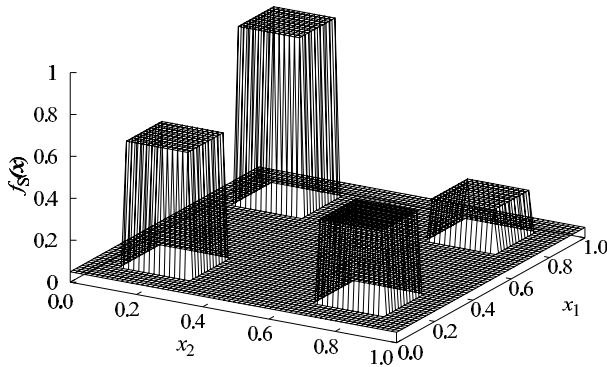


Fig. 10 $f_S(\mathbf{x})|_{\text{uniRGO}}$ defined in Table 2.

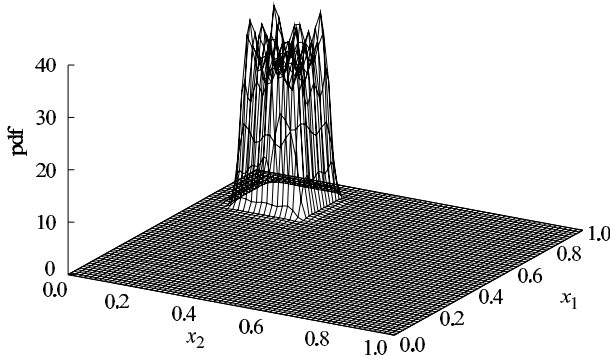


Fig. 11 $\tilde{p}_{\text{opt}}(\mathbf{x})$ for $f_S(\mathbf{x})|_{\text{uniRGO}}$ obtained using Algorithm 1.

to the global optimum ($\mathbf{x}_{\text{opt}} \stackrel{\text{def}}{=} (x_{1\text{opt}}, x_{2\text{opt}})^T$). In Fig. 9, we can see that $\tilde{p}_{\text{opt}}(\mathbf{x})$ is also a good approximation of $\delta(\mathbf{x} - \mathbf{x}_{\text{opt}})$. This result is consistent with the result in Sect. 4.2 because the $f_G(\mathbf{x})|_{\text{uniGO}}$ shown in Fig. 8 is a differentiable function and has a unique global optimum in the same manner as the function in Sect. 4.2.

In contrast, $f_S(\mathbf{x})|_{\text{uniRGO}}$ shown in Fig. 10 contains regions where a derivative cannot be derived. Moreover, the derivative in the regions where the function is differentiable is equal to zero. Therefore, it is not possible to apply gradient methods to $f_S(\mathbf{x})|_{\text{uniRGO}}$ because they use a derivative of the objective function. On the

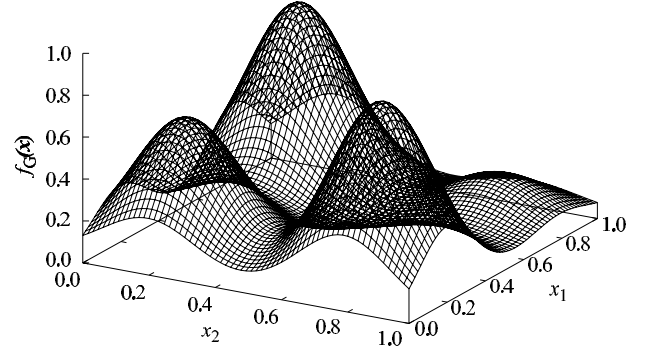


Fig. 12 $f_G(\mathbf{x})|_{\text{twoGO}}$ defined in Table 2.

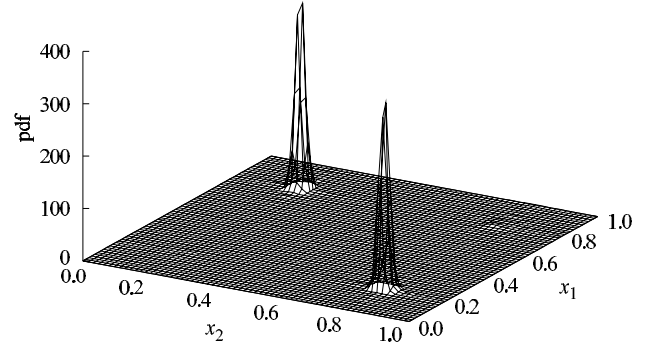


Fig. 13 $\tilde{p}_{\text{opt}}(\mathbf{x})$ for $f_G(\mathbf{x})|_{\text{twoGO}}$ obtained using Algorithm 1.

other hand, many global optimums are obtained using random search methods. Thus, a somewhat complicated operation has to be added if it is necessary to evaluate the relationship of the global optimums and to find the value representing them. In contrast, Algorithm 1 provides, without any additional operation, a solution that is the center of the global optimums to represent them. In Table 5, \mathbf{x}_{opt} and $E[\mathbf{x}]_{\text{opt}}$ for $f_S(\mathbf{x})|_{\text{uniRGO}}$ denote the center of the global optimums and its approximation obtained using Algorithm 1, respectively. We can see from Table 5 that Algorithm 1 provides a solution that represents the global optimums. The distribution of the global optimums is obtained using their higher-order statistics such as standard deviation and n th moment defined by Eqs. (23) and (24).

As shown in Fig. 12 and Table 5, $f_G(\mathbf{x})|_{\text{twoGO}}$ has two global optimums. Although an approximation of the pdf that expresses the global optimums was obtained, as shown in Fig. 13, the approximation of the global optimum in Table 5 is wrong because it is assumed in Eq. (26) that there is a unique global optimum. Deriving good approximations of the global optimums from the pdf in Fig. 13 is a subject for future study.

As in the case of $f_G(\mathbf{x})|_{\text{twoGO}}$, a correct approximation of the global optimum is not always obtained. It is thus necessary to judge whether the approximations of the global optimums in Table 5 are good or

Table 5 Global optimums and their approximations obtained using Algorithm 1.

	$E[x_1]_{\text{opt}}$	$E[x_2]_{\text{opt}}$	$x_{1\text{opt}}$	$x_{2\text{opt}}$
$f_G(\mathbf{x}) _{\text{uniGO}}$	0.799	0.202	0.8	0.2
$f_S(\mathbf{x}) _{\text{uniRGO}}$	0.798	0.201	0.8	0.2
$f_G(\mathbf{x}) _{\text{twoGO}}$	0.509	0.509	0.8(0.2)	0.2(0.8)

Table 6 Approximately maximum values of objective functions obtained using Algorithm 1.

	$E[y]_{\text{opt}}$	$f(E[\mathbf{x}]_{\text{opt}})$
$f_G(\mathbf{x}) _{\text{uniGO}}$	0.881	0.900
$f_S(\mathbf{x}) _{\text{uniRGO}}$	0.885	0.900
$f_G(\mathbf{x}) _{\text{twoGO}}$	0.886	0.077

not without having any knowledge useful for deriving the global optimum. This can be done by comparing $E[y]_{\text{opt}}$ with $f(E[\mathbf{x}]_{\text{opt}})$. As shown in Table 6, $E[y]_{\text{opt}}$ is almost equal to $f(E[\mathbf{x}]_{\text{opt}})$ when the approximation of the global optimum is good ($f_G(\mathbf{x})|_{\text{uniGO}}$ and $f_S(\mathbf{x})|_{\text{uniRGO}}$). On the other hand, $E[y]_{\text{opt}}$ is far from $f(E[\mathbf{x}]_{\text{opt}})$ when the approximation of the global optimum is bad ($f_G(\mathbf{x})|_{\text{twoGO}}$). Therefore, we can conclude that we can derive a good approximation of the global optimum if there is a unique global optimum and that we can judge whether the approximation is good or bad.

5. Conclusion

A method was developed for deriving the global optimum of a nonlinear objective function with multiple local optimums. The objective function is expanded into a linear wave coefficient equation, and the problem of maximizing the objective function is reduced to that of maximizing a quadratic function with respect to the wave coefficients. The method was examined by computer simulation for various objective functions. It was shown that a good approximation of the global optimum of each objective function can always be obtained if the objective function has a unique global optimum and that the accuracy of the approximation increases with the degree of expansion of the nonlinear objective function. Although the calculation cost of the method is not lower than that of conventional methods, the method is based on a novel idea. Since it uses the wave function of the objective function, it should be possible to extend it into a quantum computing algorithm.

References

- [1] A. V. Levy and A. Montalvo, "The tunneling algorithm for the global minimization of functions," *SIAM J. Sci. Statist. Comput.*, vol. 6, no. 1, pp. 15–29, 1985.
- [2] P. Liu and B. J. Berne, "Quantum path minimization: an efficient method for global optimization," *Journal of Chemical Physics*, vol. 118, no. 7, pp. 2999–3005, Feb. 2003.
- [3] Wenzel W. Hamacher K, "Stochastic tunneling approach for global minimization of complex potential energy landscapes," *Physical Review Letters*, vol. 82, no. 15, pp. 3003–

3007, April 1999.

- [4] H. Nozawa, "A neural network model as a globally coupled map and applications based on chaos," *Chaos*, vol. 2, no. 3, pp. 377–386, 1992.
- [5] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," *Biological Cybernetics*, vol. 52, pp. 141–152, 1985.
- [6] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines," *Cognitive Science*, vol. 9, pp. 147–169, 1985.
- [7] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.
- [8] I. Ono and S. Kobayashi, "A real-coded genetic algorithm for function optimization using unimodal normal distribution crossover," *Proc. 7th Int. Conf. on Genetic Algorithms*, pp. 246–253, 1997.
- [9] G. E. Santoro and E. Tosatti, "Optimization using quantum mechanics: quantum annealing through adiabatic evolution," *J. Phys. A: Math. Gen.*, vol. 39, no. 36, pp. R393–R431, Sept. 2006.
- [10] S. Morita and H. Nishimori, "Convergence theorems for quantum annealing," *J. Phys. A: Math. Gen.*, vol. 39, pp. 13903–13920, 2006.
- [11] H. Nishimori and Y. Nonomura, "Quantum effects in neural networks," *Journal of the Physical Society of Japan*, vol. 65, no. 12, pp. 3780–3796, 1996.
- [12] M. Akazawa and E. Tokuda and N. Asahi and Y. Amemiya, "Quantum Hopfield network using single-electron circuits—a novel Hopfield network free from the local-minimum difficulty," *Analog Integrated Circuits and Signal Processing*, vol. 24, no. 1, pp. 51–57, June 2000.
- [13] S. Gasiorowicz, *Quantum Physics 2nd Ed.*, John Wiley & Sons Inc., USA, 1996.
- [14] H. Satoh, "Approximation and analysis of nonlinear equations in a moment vector space," *IEICE Trans. Fundamentals*, vol. E89-A, no. 1, pp. 270–279, Jan. 2006.
- [15] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, UK, 2000.
- [16] I. N. Bronshtein and K. A. Semendyayev, *Handbook of Mathematics*, Springer-Verlag, UK, 1997.
- [17] D. G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, USA, 1984.
- [18] A. Papoulis, *The Fourier Integral and its Applications*, McGraw-Hill, USA, 1962.

Appendix A: Basis for Function Approximation

An orthonormal basis is summarized in this appendix. Let $h(\mathbf{k})$ be the Fourier coefficient, $\mathbf{k} \stackrel{\text{def}}{=} (k_1, \dots, k_{d_x})^T \in \mathcal{Z}$ be the index vector of the Fourier coefficient, and \mathcal{Z} be the set of \mathbf{k} that are used for the index vectors. The Fourier series expansion for function $f(\mathbf{x})$ is defined by [16]

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathcal{Z}} h(\mathbf{k}) K(\mathbf{x}, \mathbf{k}), \quad (\text{A} \cdot 1)$$

$$h(\mathbf{k}) \stackrel{\text{def}}{=} \int_{\mathcal{D}_x} f(\mathbf{x}) K^*(\mathbf{x}, \mathbf{k}) d\mathbf{x}, \quad (\text{A} \cdot 2)$$

where $\mathbf{x} \stackrel{\text{def}}{=} (x_1, \dots, x_{d_x})^T$ is the state vector of dimension d_x , $\mathcal{D}_{\mathbf{x}} \stackrel{\text{def}}{=} \{\mathbf{x} | x_{\min d} \leq x_d \leq x_{\max d}, 1 \leq d \leq d_x\}$ is the domain of the definition of \mathbf{x} , superscript $*$ denotes a complex conjugate, $\{K(\mathbf{x}, \mathbf{k})\}$ is a multi-dimensional orthonormal basis, and $K(\mathbf{x}, \mathbf{k})$ is defined by

$$K(\mathbf{x}, \mathbf{k}) \stackrel{\text{def}}{=} \prod_{d=1}^{d_x} K_d(x_d, k_d). \quad (\text{A} \cdot 3)$$

Here, $\{K_d(x_d, k_d)\}$ is a one-dimensional orthonormal basis.

Let $\{\phi_i(\cdot)\}$ be a basis the element of which is defined by

$$\phi_i(\mathbf{x}) \stackrel{\text{def}}{=} K(\mathbf{x}, \mathbf{k}), \quad (\text{A} \cdot 4)$$

where i is the index of the basis. When $\mathcal{Z}_d \stackrel{\text{def}}{=} \{0, 1, \dots, N_d\}$ and \mathcal{Z} is given by the Cartesian product as $\mathcal{Z} = \mathcal{Z}_1 \times \mathcal{Z}_2 \times \dots \times \mathcal{Z}_{d_s}$, the relationship between \mathbf{k} and i can be obtained using

$$i = \sum_{d=1}^{d_x} k_d \prod_{d'=d+1}^{d_x} (N_{d'} + 1), \quad (\text{A} \cdot 5)$$

where N_d is the degree of expansion of x_d . Let N be the degree of expansion of \mathbf{x} . When Eq. (A.5) holds, N is expressed by

$$N = \prod_{d=1}^{d_x} (N_d + 1) - 1, \quad (\text{A} \cdot 6)$$

where the dimension of the feature space with the basis is $N + 1$. The relationship between i and \mathbf{k} is referred to as the index table.

The element of the orthonormal basis based on the complex Fourier series is defined as [18]

$$K_d(x_d, k_d) \stackrel{\text{def}}{=} \begin{cases} \sqrt{\frac{1}{D_{\mathbf{x}d}}} & \text{for } k_d = 0 \\ \sqrt{\frac{1}{D_{\mathbf{x}d}}} \exp(-i \frac{k_d+1}{2} \omega_{0d}(x_d - x_{\min d})) & \text{for } k_d = 1, 3, \dots \\ \sqrt{\frac{1}{D_{\mathbf{x}d}}} \exp(i \frac{k_d}{2} \omega_{0d}(x_d - x_{\min d})) & \text{for } k_d = 2, 4, \dots \end{cases}$$

where i denotes the imaginary unit, $\omega_{0d} \stackrel{\text{def}}{=} 2\pi/D_{\mathbf{x}d}$, and $D_{\mathbf{x}d} \stackrel{\text{def}}{=} x_{\max d} - x_{\min d}$.

Appendix B: Wave Coefficient of Delta Function

Delta function $\delta(\mathbf{x} - \hat{\mathbf{x}})$ is defined by [13]

$$\delta(\mathbf{x} - \hat{\mathbf{x}}) \stackrel{\text{def}}{=} \lim_{a \rightarrow 0} \prod_{d=1}^{d_x} \frac{1}{a\sqrt{\pi}} \exp(-\frac{(x_d - \hat{x}_d)^2}{a^2}), \quad (\text{A} \cdot 7)$$

where $\hat{\mathbf{x}} \stackrel{\text{def}}{=} (\hat{x}_1, \dots, \hat{x}_{d_x})^T$ is a constant vector. Consider the pdf defined by $p(\mathbf{x}) \stackrel{\text{def}}{=} \delta(\mathbf{x} - \hat{\mathbf{x}})$. Its wave function is expressed by [13]

$$\Psi(\mathbf{x}, \hat{\mathbf{x}}) = \lim_{a \rightarrow 0} \prod_{d=1}^{d_x} \frac{1}{(a^2\pi)^{1/4}} \exp(-\frac{(x_d - \hat{x}_d)^2}{2a^2} + i\kappa_0(x_d - \hat{x}_d)), \quad (\text{A} \cdot 8)$$

where κ_0 is the angular wave number.

Let $\hat{\mathbf{c}} \stackrel{\text{def}}{=} (\hat{c}_0, \dots, \hat{c}_{N_x})^T$ be the normalized wave coefficient vector of $\Psi(\mathbf{x}, \hat{\mathbf{x}})$. Using Eq. (A.2), we obtain

$$\hat{c}_i = \int \Psi(\mathbf{x}, \hat{\mathbf{x}}) \psi_i^*(\mathbf{x}) d\mathbf{x}.$$

Because $\Psi(\mathbf{x}, \hat{\mathbf{x}})$ in the above equation is independent of i , there is an appropriate constant, ξ , and substitution of Eq. (A.8) into the above equation yields

$$\hat{c}_i = \xi^{-1} \psi_i^*(\hat{\mathbf{x}}). \quad (\text{A} \cdot 9)$$

Because \hat{c}_i is a normalized wave coefficient, the following equation should hold:

$$\begin{aligned} \sum_{i=0}^{N_x} \hat{c}_i \hat{c}_i^* &= \xi^{-2} \sum_{i=0}^{N_x} \psi_i^*(\hat{\mathbf{x}}) \psi_i(\hat{\mathbf{x}}) \\ &= \xi^{-2} (N_x + 1) \prod_{d=1}^{d_x} D_{\mathbf{x}d}^{-1} \\ &= 1. \end{aligned}$$

We thus obtain

$$\xi = \sqrt{N_x + 1} \prod_{d=1}^{d_x} \frac{1}{\sqrt{D_{\mathbf{x}d}}}. \quad (\text{A} \cdot 10)$$

Appendix C: Wave Coefficient Equation for Arbitrary Probability Density Function

Because the WCE in Eq. (14) was derived assuming that the pdfs of \mathbf{x} and y are delta functions, it is not clear whether Eq. (14) holds for any wave coefficient vector. In this appendix, Eq. (14) is extended to a WCE for arbitrary wave coefficient vectors.

In the same manner as for Eq. (10) in Sect. 2.2, consider the following wave coefficient vector for $\hat{\mathbf{x}}_\ell \in \mathcal{D}_{\mathbf{x}}$:

$$\hat{\mathbf{q}}_\ell \stackrel{\text{def}}{=} \xi_q^{-1} \psi^*(\hat{\mathbf{x}}_\ell), \quad (\text{A} \cdot 11)$$

where $\ell = 1, 2, \dots, L$, $\hat{\mathbf{x}}_\ell \neq \hat{\mathbf{x}}_m$ for $\ell \neq m$, and ξ_q is defined in the same manner as in Eq. (A.10). Let $p_\ell(\mathbf{x})$ be the pdf corresponding to $\hat{\mathbf{q}}_\ell$. Using Eq. (12), we can express $p_\ell(\mathbf{x})$ as

$$p_\ell(\mathbf{x}) \cong \|\hat{\mathbf{q}}_\ell\|^{-2} \psi^T(\mathbf{x}) \hat{\mathbf{q}}_\ell \hat{\mathbf{q}}_\ell^\dagger \psi^*(\mathbf{x}). \quad (\text{A} \cdot 12)$$

Consider pdf $p(\mathbf{x})$ defined by

$$p(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{\ell=1}^L w_{\ell} p_{\ell}(\mathbf{x}), \quad (\text{A} \cdot 13)$$

where $w_{\ell} \in \mathcal{R}$ and $w_{\ell} \geq 0$. Consider \hat{y}_{ℓ} obtained using Eq. (1) as

$$\hat{y}_{\ell} = f(\hat{\mathbf{x}}_{\ell}).$$

When $p(\mathbf{x})$ is defined by Eq. (A·13) and y is obtained by Eq. (1), the probability density function of y is expressed as

$$p(y) \cong \sum_{\ell=1}^L w_{\ell} p_{\ell}(y), \quad (\text{A} \cdot 14)$$

where $p_{\ell}(y)$ is defined by

$$p_{\ell}(y) \stackrel{\text{def}}{=} \|\hat{\mathbf{r}}_{\ell}\|^{-2} \boldsymbol{\psi}^T(y) \hat{\mathbf{r}}_{\ell} \hat{\mathbf{r}}_{\ell}^{\dagger} \boldsymbol{\psi}^*(y), \quad (\text{A} \cdot 15)$$

$$\hat{\mathbf{r}}_{\ell} \stackrel{\text{def}}{=} \xi_r^{-1} \boldsymbol{\psi}^*(\hat{y}_{\ell}), \quad (\text{A} \cdot 16)$$

in the same manner as for Eq. (13). This is clear because $p_{\ell}(\mathbf{x}) \cong \delta(\mathbf{x} - \hat{\mathbf{x}}_{\ell})$ and $p_{\ell}(y) \cong \delta(y - \hat{y}_{\ell})$.

When Eq. (A·14) holds, Eq. (14) can be rewritten as

$$\hat{\mathbf{r}}_{\ell} = A \hat{\mathbf{q}}_{\ell}. \quad (\text{A} \cdot 17)$$

Consider wave coefficient vector \mathbf{q} defined by

$$\mathbf{q} \stackrel{\text{def}}{=} \sum_{\ell=1}^L w'_{\ell} \hat{\mathbf{q}}_{\ell}. \quad (\text{A} \cdot 18)$$

We can express an arbitrary \mathbf{q} by using a sufficiently large L , appropriate constants $w'_{\ell} \in \mathcal{R}$, and an appropriate $\hat{\mathbf{q}}_{\ell}$. Consider wave coefficient vector \mathbf{r} defined by

$$\mathbf{r} \stackrel{\text{def}}{=} \sum_{\ell=1}^L w'_{\ell} \hat{\mathbf{r}}_{\ell}. \quad (\text{A} \cdot 19)$$

By substituting Eqs. (A·18) and (A·19) into Eq. (A·17), we obtain

$$\mathbf{r} = A \mathbf{q}. \quad (\text{A} \cdot 20)$$

Let us define the relationship between w_{ℓ} and w'_{ℓ} as

$$w_{\ell} = \left\| \sum_m w'_m \mathbf{q}_m \right\|^{-2} \|w'_{\ell} \hat{\mathbf{q}}_{\ell}\|^{-2}. \quad (\text{A} \cdot 21)$$

Consider wave functions $\Psi(\mathbf{x})$ and $\Psi(y)$ defined by

$$\Psi(\mathbf{x}) \stackrel{\text{def}}{=} \|\mathbf{q}\|^{-1} \boldsymbol{\psi}^T(\mathbf{x}) \mathbf{q}, \quad (\text{A} \cdot 22)$$

$$\Psi(y) \stackrel{\text{def}}{=} \|\mathbf{r}\|^{-1} \boldsymbol{\psi}^T(y) \mathbf{r}. \quad (\text{A} \cdot 23)$$

Because $\|\hat{\mathbf{r}}_{\ell}\| \cong 1$ when we set $\|\hat{\mathbf{q}}_{\ell}\| = 1$, $p(\mathbf{x})$ in Eq.

(A·13) and $p(y)$ in Eq. (A·14) can be expressed as

$$p(\mathbf{x}) \cong \|\mathbf{q}\|^{-2} \boldsymbol{\psi}^T(\mathbf{x}) \mathbf{q} \mathbf{q}^{\dagger} \boldsymbol{\psi}^*(\mathbf{x}), \quad (\text{A} \cdot 24)$$

$$p(y) \cong \|\mathbf{r}\|^{-2} \boldsymbol{\psi}^T(y) \mathbf{r} \mathbf{r}^{\dagger} \boldsymbol{\psi}^*(y). \quad (\text{A} \cdot 25)$$

As described above, Eq. (A·20) shows that we can derive \mathbf{r} for an arbitrary \mathbf{q} using matrix A , and Eqs. (A·24) and (A·25) show that $p(\mathbf{x})$ and $p(y)$ related to $p(\mathbf{x})$ by Eq. (1) can be obtained using \mathbf{q} and \mathbf{r} . Thus, Eq. (A·20) expresses the nonlinear function in Eq. (1) for an arbitrary wave coefficient vector.

Appendix D: Optimization in Wave Coefficient Space

The relationship between the optimization problem for y and that for $E[y]$ is described here. First, we clarify the meaning of maximizing $E[y]$. Let y_{\max} be the maximum value of $y = f(\mathbf{x})$, and recall that

- (1) $E[y]$ is defined as $E[y] \stackrel{\text{def}}{=} \int y p(y) dy$, as in Eq. (18).
- (2) The definitions of A , \mathbf{q} , and Y and Eqs. (15) and (17) mean that an arbitrary $p(y)$ including $\delta(y - y_{\max})$ can be obtained by modifying \mathbf{q} .

From the above definitions, maximizing $E[y]$ denotes adjusting \mathbf{q} so as to maximize $E[y]$.

Next, recall that y_{\max} is equal to $E[y]$ if $p(y) = \delta(y - y_{\max})$. This means that we can obtain $E[y]$ such that y_{\max} is equal to $E[y]$ by adjusting the shape of $p(y)$. Because we can obtain an arbitrary shape of $p(y)$ by adjusting \mathbf{q} , it is possible to make $E[y]$ equal y_{\max} by adjusting \mathbf{q} .

In contrast, if $\max E[y] = y_{\max}$, $p(y) = \delta(y - y_{\max})$. Because an arbitrary $p(y)$ including $\delta(y - y_{\max})$ can be obtained by modifying \mathbf{q} , there exists \mathbf{q} such that $\max E[y] = y_{\max}$. Therefore, we can replace the optimization problem for y with that for $E[y]$.

The change in $p(y)$ is shown in Fig. 4, and we can confirm that $p(y) = \delta(y - y_{\max})$ when $E[y] = y_{\max}$. Note that the $\hat{p}_{\text{opt}}(\mathbf{x})$ that provides y_{\max} is not always a delta function, as shown in Figs. 11 and 13, and that it was confirmed that they also provided $p(y) = \delta(y - y_{\max})$ although their figures were omitted in the paper.